

Результаты моделирования могут быть использованы для понимания происхождения и эволюции систем многих тел.

Дисперсия, а.е.<sup>2</sup>

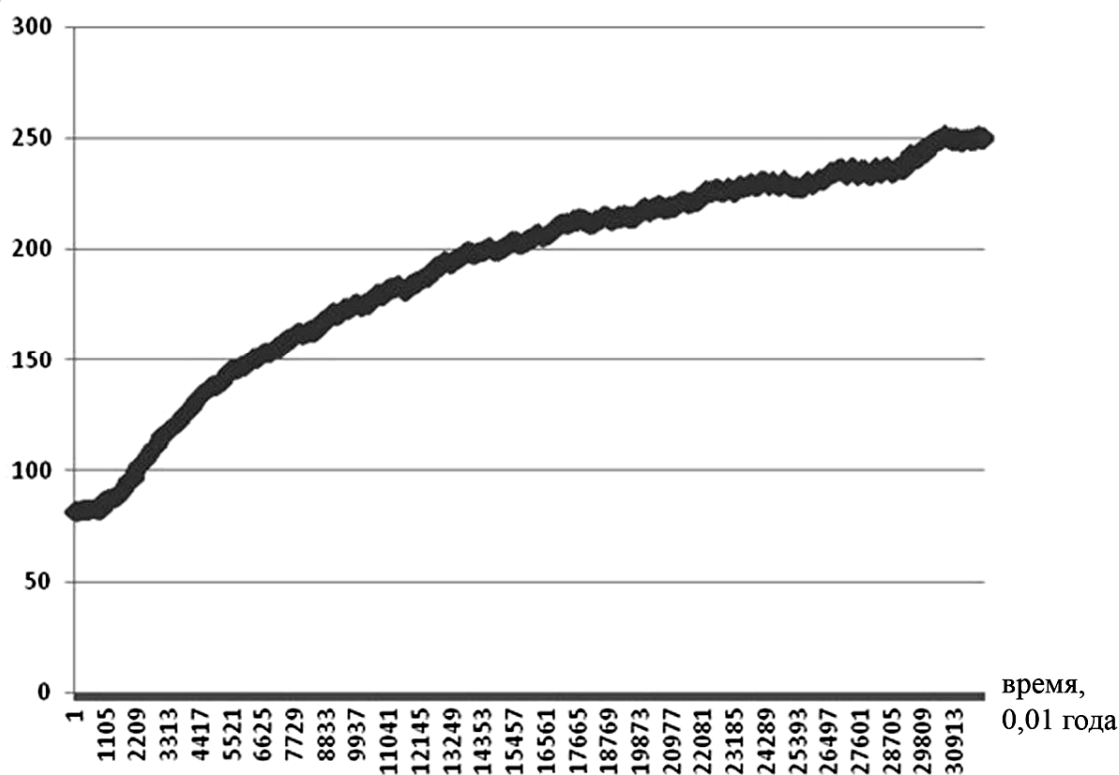


Рис. 1. Зависимость дисперсии функции распределения по модулю радиус-вектора от времени системы с 20ю планетами, расположенными равномерно в интервале от 1 до 30 а.е. от центрального тела

## ИНТЕНТ-АНАЛИЗ КАК КОМПЬЮТЕРНЫЙ МЕТОД ОБРАБОТКИ ТЕКСТА

Минюров Е.С.

Уральский федеральный университет имени первого Президента России  
Б.Н. Ельцина, г. Екатеринбург, Россия

Бурный рост текстовой информации актуализировал необходимость развития компьютерной лингвистики, область приложений которой постоянно расширяется. Особенно интенсивно ведутся исследования в направлении Text Mining – извлечение информации из текстов, что требуется при решении задач экономической и производственной аналитики. К этому направлению относятся две другие близкие задачи: выделение мнений (Opinion Mining) и оценка тональности текстов (Sentiment Analysis). Одним из перспективных направлений компьютерных исследований текстов является разработка автоматизированного

варианта контент-анализа, примером которого могут служить пакеты «ЛЕКТА» и «БААЛ».

Все более актуальной прикладной задачей становится извлечение из текстов не только прямой информации, но и скрытых намерений. За рубежом исследования в этом направлении ведутся очень активно. Так, в мае 2012 американская компания DARPA объявила о начале разработки средства интент-анализа текстов DEFT. Компания стремится создать технологию, которая может снимать фильтры, затемняющие смысл текста, а также выполнять отбраковку ложной информации. Чтобы быть успешной, технология DEFT будет создана на существующих программах DARPA и академических исследованиях в области глубокого понимания языка и искусственного интеллекта. В России пока нет таких систем обработки текстов. В этой связи, проблема исследования и разработки отечественных аналогов интент-анализатора текста имеет высокий уровень актуальности [1].

Выдвигая задачу разработки автоматизированного варианта интент-анализа, мы считаем важным опираться в качестве ключевого на понятие «система анализа текста». Это специально разработанная компьютерная система, которая должна иметь возможность проанализировать текст, поданный пользователем на вход, с точки зрения синтаксиса (структуры предложений), семантики (понятий, применяемых в тексте) и прагматики (правильности употребления понятий и целей их употребления). Система должна сгенерировать свой отклик во внутреннем представлении, пригодном для логического вывода, и просинтезировать свой отклик на естественном языке [2]. Ключевыми для интент-анализа, на наш взгляд, являются постсинтаксический анализ (уточнение смысла, заложенного в слова и выраженного при помощи различных средств языка: предлогов, префиксов или аффиксов, создающих ту или иную словоформу) и семантический анализ (определение смысла фраз, выявление подтекста).

Мы считаем, что построение алгоритма автоматизированного варианта интент-анализа для обнаружения в тексте скрытых намерений может опираться на анализ подтекста как результата межуровневых связей, соединения эксплицитной и имплицитной информации, заложенной в тексте. Исследователи выделяют два вида средств выражения подтекста: собственно языковые средства, обладающие способностью выражать оба вида информации, и приемы их использования, которые являются средством дополнительного маркирования, «переключения» их с функции выражения основной информации на функцию выражения дополнительной информации [3]. Основные типы этих маркеров - нарушение стандартного функционирования языковых единиц и их избыточное использование.

1. <http://www.darpa.mil/NewsEvents/Releases/2012/05/03a.aspx>
2. Большакова Е.И., Клышинский Э.С. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика, М.: МИЭМ (2011).
3. Масленникова, А. А. Лингвистическая интерпретация скрытых смыслов, СПб: Союз (1999).